Dynamical Systems Gradient method for solving nonlinear equations with monotone operators

N. S. Hoang†* A. G. Ramm†‡

†Mathematics Department, Kansas State University,
Manhattan, KS 66506-2602, USA

Abstract

A version of the Dynamical Systems Gradient Method for solving ill-posed nonlinear monotone operator equations is studied in this paper. A discrepancy principle is proposed and justified. A numerical experiment was carried out with the new stopping rule. Numerical experiments show that the proposed stopping rule is efficient. Equations with monotone operators are of interest in many applications.

Mathematics Subject Classification. 47J05, 47J06, 47J35, 65R30
Keywords. Dynamical systems method (DSM), nonlinear operator equations, monotone operators, discrepancy principle.

1 Introduction

In this paper we study a version of the Dynamical Systems Method (DSM) (see [10]) for solving the equation

$$F(u) = f, \tag{1}$$

where F is a nonlinear, twice Fréchet differentiable, monotone operator in a real Hilbert space H, and equation (1) is assumed solvable, possibly nonuniquely. Monotonicity means that

$$\langle F(u) - F(v), u - v\rangle \geq 0, \quad \forall u, v \in H. \tag{2}$$

Equations with monotone operators are important in many applications and were studied extensively, see, for example, [5], [7], [21], [24], and references therein. One encounters many technical and physical problems with such operators in the cases where dissipation of energy occurs. For example, in [9] and [8], Chapter 3, pp.156-189, a wide class of nonlinear dissipative systems is studied, and the basic equations of such systems can be reduced to equation (1) with monotone operators. Numerous examples of equations with monotone operators can be found in [5] and references mentioned above. In [19] and [20] it is proved

*Email: nguyenhs@math.ksu.edu
‡Corresponding author. Email: ramm@math.ksu.edu

that any solvable linear operator equation with a closed, densely defined operator in a Hilbert space H can be reduced to an equation with a monotone operator and solved by a convergent iterative process.

In this paper, apparently for the first time, the convergence of the Dynamical Systems Gradient method is proved under natural assumptions and convergence of a corresponding iterative method is established. No special assumptions of smallness of the nonlinearity or other special properties of the nonlinearity are imposed. No source-type assumptions are used. Consequently, our result is quite general and widely applicable. It is well known, that without extra assumptions, usually, source-type assumption about the right-hand side, or some assumption concerning the smoothness of the solution, one cannot get a specific rate of convergence even for linear ill-posed equations (see, for example, [10], where one can find a proof of this statement). On the other hand, such assumptions are often difficult to verify and often they do not hold. By this reason we do not make such assumptions.

The result of this paper is useful both because of its many possible applications and because of its general nature. Our novel technique consists of an application of some new inequalities. Our main results are formulated in Theorems 17 and 19, and also in several lemmas, for example, in Lemmas 3, 4, 8, 9, 11, 12. Lemmas 3, 4, 11, 12 may be useful in many other problems.

In [23] a stationary equation $F(u) = f$ with a nonlinear monotone operator F was studied. The assumptions A1-A3 on p.197 in [23] are more restrictive than ours, and the Rule R2 on p.199, formula (4.1) in [23] for the choice of the regularization parameter is quite different from our rule and is more difficult to use it computationally: one has to solve a nonlinear equation (equation (4.1) in [23]) in order to find the regularization parameter. To use this equation one has to invert an ill-conditioned linear operator $A+\alpha I$ for small values of α. Assumption A1 in [23] is not verifiable practically, because the solution $x^\dagger$ is not known. Assumption A3 in [23] requires F to be constant in a ball $B_r(x^\dagger)$ if $F'(x^\dagger) = 0$. Our method does not require these assumptions, and, in contrast to equation (4.1) in [23], it does not require inversion of ill-conditioned linear operators and solving nonlinear equations for finding the regularization parameter. The stopping time is chosen numerically in our method without extra computational effort by a discrepancy-type principle formulated and justified in Theorem 17, in Section 3. We give a convergent iterative process for stable solution of equation (1.1) and a stopping rule for this process.

In [23] the "source-type assumption" is made, that is, it is assumed that the right-hand side of the equation $F(u) = f$ belongs to the range of a suitable operator. This usually allows one to get some convergence rate. In our paper, as was already mentioned above, such an assumption is not used because, on the one hand, numerically it is difficult to verify such an assumption, and, on the other hand, such an assumption may be not satisfied in many cases, even in linear ill-posed problems, for example, in the case when the solution does not have extra smoothness.

We assume the nonlinearity to be twice locally Fréchet differentiable. This assumption, as we mention below, does not restrict the global growth of the nonlinearity. In many practical and theoretical problems the nonlinearities are smooth and given analytically. In these cases one can calculate F' analytically. This is the case in the example, considered in Section 4. This example is a simple model problem for non-linear Wiener-type filtering (see

[18]). If one drops the nonlinear cubic term in the equation $Bu + u^3 = f$ of this example, then the resulting equation $Bu = f$ does not have integrable solutions, in general, even for very smooth f, for example, for $f \in C^\infty([0,1])$, as shown in [18]. It is, therefore, of special interest to solve this equation numerically.

It is known (see, e.g., [10]), that the set $\mathcal{N} := \{u : F(u) = f\}$ is closed and convex if F is monotone and continuous. A closed and convex set in a Hilbert space has a unique minimal-norm element. This element in $\mathcal{N}$ we denote by y, $F(y) = f$. We assume that

$$\sup_{\|u-u_0\|\le R} \|F^{(j)}(u)\| \le M_j(R), \quad 0 \le j \le 2, \tag{3}$$

where $u_0 \in H$ is an element of H, $R > 0$ is arbitrary, and $f = F(y)$ is not known but f_δ, the noisy data, are known, and $\|f_\delta - f\| \le \delta$. Assumption (3) simplifies our arguments and does not restrict the global growth of the nonlinearity. In [12] this assumption is weakened to hemicontinuity in the problems related to the existence of the global solutions of the equations, generated by the DSM. In many applications the nonlinearity F is given analytically, and then one can calculate $F'(u)$ analytically.

If $F'(u)$ is not boundedly invertible then solving equation (1) for u given noisy data f_δ is often (but not always) an ill-posed problem. When F is a linear bounded operator many methods for stable solving of (1) were proposed (see [2], [4]–[10] and references therein). However, when F is nonlinear then the theory is less complete.

DSM consists of finding a nonlinear map $\Phi(t,u)$ such that the Cauchy problem

$$\dot{u} = \Phi(t,u), \qquad u(0) = u_0,$$

has a unique solution for all $t \ge 0$, there exists $\lim_{t\to\infty} u(t) := u(\infty)$, and $F(u(\infty)) = f$,

$$\exists!\ u(t) \quad \forall t \ge 0; \qquad \exists u(\infty); \qquad F(u(\infty)) = f. \tag{4}$$

Various choices of Φ were proposed in [10] for (4) to hold. Each such choice yields a version of the DSM.

The DSM for solving equation (1) was extensively studied in [10]–[17]. In [10], the following version of the DSM was investigated for monotone operators F:

$$\dot{u}_\delta = -\big(F'(u_\delta) + a(t)I\big)^{-1}\big(F(u_\delta) + a(t)u_\delta - f_\delta\big), \quad u_\delta(0) = u_0. \tag{5}$$

Here I denotes the identity operator in H. The convergence of this method was justified with some *a priori* choice of stopping rule.

In [22] a continuous gradient method for solving equation (1) was studied. A stopping rule of discrepancy type was introduced and justified under the assumption that F satisfies the following condition:

$$\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)\| = \eta\|F(x) - F(\tilde{x})\|, \qquad \eta < 1, \tag{6}$$

for all x, $\tilde{x}$ in some ball $B(x_0,R) \subset H$. This very restrictive assumption is not satisfied even for monotone operators. Indeed, if $F'(x) = 0$ for some $x \in B(x_0)$ then (6) implies $F(x) = f$ for all $x \in B(x_0,R)$, provided that $B(x_0,R)$ contains a solution of (1).

In this paper we consider a gradient-type version of the DSM for solving equation (1):

$$\dot{u}_\delta = -\big(F'(u_\delta)^* + a(t)I\big)\big(F(u_\delta) + a(t)u_\delta - f_\delta\big), \quad u_\delta(0) = u_0, \tag{7}$$

where F is a monotone operator and A^* denotes the adjoint to a linear operator A. If F is monotone then $F'(\cdot) := A \geq 0$. If a bounded linear operator A is defined on all of the complex Hilbert space H and $A \geq 0$, i.e., $\langle Au, u\rangle \geq 0$, $\forall u \in H$, then $A = A^*$, so A is selfadjoint. In a real Hilbert space H a bounded linear operator defined on all of H and satisfying the inequality $\langle Au, u\rangle \geq 0$, $\forall u \in H$ is not necessary selfadjoint. Example: $H = \mathbb{R}^2$, $A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$, $\langle Au, u\rangle = 2u_1^2 + u_1u_2 + u_2^2 \geq 0$, but $A^* = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix} \neq A$.

The convergence of the method (7) for any initial value u_0 is proved for a stopping rule based on a discrepancy principle. This *a posteriori* choice of stopping time t_δ is justified provided that $a(t)$ is suitably chosen.

The advantage of method (7), a modified version of the gradient method, over the Gauss-Newton method and the version (5) of the DSM is the following: no inversion of matrices is needed in (7). Although the convergence rate of the DSM (7) maybe slower than that of the DSM (5), the DSM (7) might be faster than the DSM (5) for large-scale systems due to its lower computation cost at each iteration.

In this paper we investigate a stopping rule based on a discrepancy principle (DP) for the DSM (7). The main results of this paper are Theorem 17 and Theorem 19 in which a DP is formulated, the existence of a stopping time t_δ is proved, and the convergence of the DSM with the proposed DP is justified under some natural assumptions.

2 Auxiliary results

The inner product in H is denoted $\langle u, v\rangle$. Let us consider the following equation

$$F(V_\delta) + aV_\delta - f_\delta = 0, \qquad a > 0, \tag{8}$$

where $a = const$. It is known (see, e.g., [10], [25]) that equation (8) with monotone continuous operator F has a unique solution for any $f_\delta \in H$.

Let us recall the following result from [10]:

Lemma 1 *Assume that equation* (1) *is solvable, y is its minimal-norm solution, assumptions* (2) *holds, and F is continuous. Then*

$$\lim_{a\to 0} \|V_a - y\| = 0,$$

where V_a solves (8) *with $\delta = 0$.*

Of course, under our assumption (3), F is continuous.

Lemma 2 *If* (2) *holds and F is continuous, then $\|V_\delta\| = O(\frac{1}{a})$ as $a \to \infty$, and*

$$\lim_{a\to\infty} \|F(V_\delta) - f_\delta\| = \|F(0) - f_\delta\|. \tag{9}$$

Proof. Rewrite (8) as

$$F(V_\delta) - F(0) + aV_\delta + F(0) - f_\delta = 0.$$

Multiply this equation by V_δ, use inequality $\langle F(V_\delta) - F(0), V_\delta - 0\rangle \geq 0$ and get:

$$a\|V_\delta\|^2 \leq \|f_\delta - F(0)\|\|V_\delta\|.$$

Therefore, $\|V_\delta\| = O(\frac{1}{a})$. This and the continuity of F imply (9). $\Box$

Let $a = a(t)$ be strictly monotonically decaying continuous positive function on $[0, \infty)$, $0 < a(t) \searrow 0$, and assume $a \in C^1[0, \infty)$. These assumptions hold throughout the paper and often are not repeated. Then the solution V_δ of (8) is a function of t, $V_\delta = V_\delta(t)$. From the triangle inequality one gets:

$$\|F(V_\delta(0)) - f_\delta\| \geq \|F(0) - f_\delta\| - \|F(V_\delta(0)) - F(0)\|.$$

From Lemma 2 it follows that for large $a(0)$ one has:

$$\|F(V_\delta(0)) - F(0)\| \leq M_1\|V_\delta(0)\| = O\bigg(\frac{1}{a(0)}\bigg).$$

Therefore, if $\|F(0) - f_\delta\| > C\delta$, then $\|F(V_\delta(0)) - f_\delta\| \geq (C - \epsilon)\delta$, where $\epsilon > 0$ is sufficiently small and $a(0) > 0$ is sufficiently large.

Below the words decreasing and increasing mean strictly decreasing and strictly increasing.

Lemma 3 *Assume $\|F(0) - f_\delta\| > 0$. Let $0 < a(t) \searrow 0$, and F be monotone. Denote*

$$\psi(t) := \|V_\delta(t)\|, \qquad \phi(t) := a(t)\psi(t) = \|F(V_\delta(t)) - f_\delta\|,$$

where $V_\delta(t)$ solves (8) with $a = a(t)$. Then $\phi(t)$ is decreasing, and $\psi(t)$ is increasing.

Proof. Since $\|F(0) - f_\delta\| > 0$, one has $\psi(t) \neq 0$, $\forall t \geq 0$. Indeed, if $\psi(t)\big|_{t=\tau} = 0$, then $V_\delta(\tau) = 0$, and equation (8) implies $\|F(0) - f_\delta\| = 0$, which is a contradiction. Note that $\phi(t) = a(t)\|V_\delta(t)\|$. One has

$$\begin{aligned}
0 &\leq \langle F(V_\delta(t_1)) - F(V_\delta(t_2)), V_\delta(t_1) - V_\delta(t_2)\rangle \\
&= \langle -a(t_1)V_\delta(t_1) + a(t_2)V_\delta(t_2), V_\delta(t_1) - V_\delta(t_2)\rangle \\
&= (a(t_1) + a(t_2))\langle V_\delta(t_1), V_\delta(t_2)\rangle - a(t_1)\|V_\delta(t_1)\|^2 - a(t_2)\|V_\delta(t_2)\|^2.
\end{aligned} \tag{10}$$

Thus,

$$\begin{aligned}
0 &\leq (a(t_1) + a(t_2))\|V_\delta(t_1)\|\|V_\delta(t_2)\| - a(t_1)\|V_\delta(t_1)\|^2 - a(t_2)\|V_\delta(t_2)\|^2 \\
&= (a(t_1)\|V_\delta(t_1)\| - a(t_2)\|V_\delta(t_2)\|)(\|V_\delta(t_2)\| - \|V_\delta(t_1)\|) \\
&= (\phi(t_1) - \phi(t_2))(\psi(t_2) - \psi(t_1)).
\end{aligned} \tag{11}$$

If $\psi(t_2) > \psi(t_1)$ then (11) implies $\phi(t_1) \ge \phi(t_2)$, so

$$a(t_1)\psi(t_1) \ge a(t_2)\psi(t_2) > a(t_2)\psi(t_1).$$

Thus, if $\psi(t_2) > \psi(t_1)$ then $a(t_2) < a(t_1)$ and, therefore, $t_2 > t_1$, because $a(t)$ is strictly decreasing.

Similarly, if $\psi(t_2) < \psi(t_1)$ then $\phi(t_1) \le \phi(t_2)$. This implies $a(t_2) > a(t_1)$, so $t_2 < t_1$.

Suppose $\psi(t_1) = \psi(t_2)$, i.e., $\|V_\delta(t_1)\| = \|V_\delta(t_2)\|$. From (10), one has

$$\|V_\delta(t_1)\|^2 \le \langle V_\delta(t_1), V_\delta(t_2)\rangle \le \|V_\delta(t_1)\|\|V_\delta(t_2)\| = \|V_\delta(t_1)\|^2.$$

This implies $V_\delta(t_1) = V_\delta(t_2)$, and then equation (8) implies $a(t_1) = a(t_2)$. Hence, $t_1 = t_2$, because $a(t)$ is strictly decreasing.

Therefore $\phi(t)$ is decreasing and $\psi(t)$ is increasing. □

Lemma 4 *Suppose that $\|F(0) - f_\delta\| > C\delta$, $C > 1$, and $a(0)$ is sufficiently large. Then, there exists a unique $t_1 > 0$ such that $\|F(V_\delta(t_1)) - f_\delta\| = C\delta$.*

Proof. The uniqueness of t_1 follows from Lemma 3 because $\|F(V_\delta(t)) - f_\delta\| = \phi(t)$, and ϕ is decreasing. We have $F(y) = f$, and

$$\begin{aligned}
0 &= \langle F(V_\delta) + aV_\delta - f_\delta, F(V_\delta) - f_\delta\rangle \\
&= \|F(V_\delta) - f_\delta\|^2 + a\langle V_\delta - y, F(V_\delta) - f_\delta\rangle + a\langle y, F(V_\delta) - f_\delta\rangle \\
&= \|F(V_\delta) - f_\delta\|^2 + a\langle V_\delta - y, F(V_\delta) - F(y)\rangle + a\langle V_\delta - y, f - f_\delta\rangle + a\langle y, F(V_\delta) - f_\delta\rangle \\
&\ge \|F(V_\delta) - f_\delta\|^2 + a\langle V_\delta - y, f - f_\delta\rangle + a\langle y, F(V_\delta) - f_\delta\rangle.
\end{aligned}$$

Here the inequality $\langle V_\delta - y, F(V_\delta) - F(y)\rangle \ge 0$ was used. Therefore

$$\begin{aligned}
\|F(V_\delta) - f_\delta\|^2 &\le -a\langle V_\delta - y, f - f_\delta\rangle - a\langle y, F(V_\delta) - f_\delta\rangle \\
&\le a\|V_\delta - y\|\|f - f_\delta\| + a\|y\|\|F(V_\delta) - f_\delta\| \\
&\le a\delta\|V_\delta - y\| + a\|y\|\|F(V_\delta) - f_\delta\|.
\end{aligned} \tag{12}$$

On the other hand, we have

$$\begin{aligned}
0 &= \langle F(V_\delta) - F(y) + aV_\delta + f - f_\delta, V_\delta - y\rangle \\
&= \langle F(V_\delta) - F(y), V_\delta - y\rangle + a\|V_\delta - y\|^2 + a\langle y, V_\delta - y\rangle + \langle f - f_\delta, V_\delta - y\rangle \\
&\ge a\|V_\delta - y\|^2 + a\langle y, V_\delta - y\rangle + \langle f - f_\delta, V_\delta - y\rangle,
\end{aligned}$$

where the inequality $\langle V_\delta - y, F(V_\delta) - F(y)\rangle \ge 0$ was used. Therefore,

$$a\|V_\delta - y\|^2 \le a\|y\|\|V_\delta - y\| + \delta\|V_\delta - y\|.$$

This implies

$$a\|V_\delta - y\| \le a\|y\| + \delta. \tag{13}$$

From (12) and (13), and an elementary inequality $ab \le \epsilon a^2 + \frac{b^2}{4\epsilon}$, $\forall \epsilon > 0$, one gets:

$$\begin{aligned} \|F(V_\delta) - f_\delta\|^2 &\le \delta^2 + a\|y\|\delta + a\|y\|\|F(V_\delta) - f_\delta\| \\ &\le \delta^2 + a\|y\|\delta + \epsilon\|F(V_\delta) - f_\delta\|^2 + \frac{1}{4\epsilon}a^2\|y\|^2, \end{aligned} \tag{14}$$

where $\epsilon > 0$ is fixed, independent of t, and can be chosen arbitrary small. Let $t \to \infty$ and $a = a(t) \searrow 0$. Then (14) implies

$$\overline{\lim}_{t\to\infty}(1-\epsilon)\|F(V_\delta) - f_\delta\|^2 \le \delta^2.$$

This, the continuity of F, the continuity of $V_\delta(t)$ on $[0, \infty)$, and the assumption $\|F(0) - f_\delta\| > C\delta$ imply that equation $\|F(V_\delta(t)) - f_\delta\| = C\delta$ must have a solution $t_1 > 0$. The uniqueness of this solution has already established. $\square$

Remark 5 From the proof of Lemma 4 one obtains the following claim:
If $t_n \nearrow \infty$ then there exists a unique $n_1 > 0$ such that

$$\|F(V_{n_1+1}) - f_\delta\| \le C\delta < \|F(V_{n_1}) - f_\delta\|, \qquad V_n := V_\delta(t_n).$$

Remark 6 From Lemma 2 and Lemma 3 one concludes that

$$a_n\|V_n\| = \|F(V_n) - f_\delta\| \le \|F(0) - f_\delta\|, \qquad a_n := a(t_n), \quad \forall n \ge 0.$$

Remark 7 Let $V := V_\delta(t)|_{\delta=0}$, so

$$F(V) + a(t)V - f = 0.$$

Let y be the minimal-norm solution to equation (1). We claim that

$$\|V_\delta - V\| \le \frac{\delta}{a}. \tag{15}$$

Indeed, from (8) one gets

$$F(V_\delta) - F(V) + a(V_\delta - V) = f - f_\delta.$$

Multiply this equality with $(V_\delta - V)$ and use the monotonicity of F to get

$$a\|V_\delta - V\|^2 \le \delta\|V_\delta - V\|.$$

This implies (15). Similarly, multiplying the equation

$$F(V) + aV - F(y) = 0,$$

by $V - y$ one derives the inequality:

$$\|V\| \le \|y\|. \tag{16}$$

Similar arguments one can find in [10].

From (15) and (16), one gets the following estimate:

$$\|V_\delta\| \le \|V\| + \frac{\delta}{a} \le \|y\| + \frac{\delta}{a}. \tag{17}$$

Lemma 8 *Suppose* $a(t) = \frac{d}{(c+t)^b}$, $\varphi(t) = \int_0^t \frac{a^2(s)}{2} ds$ *where* $b \in (0, \frac{1}{4}]$, d *and* c *are positive constants. Then*

$$\frac{d^2}{2}\left(1 - \frac{2b}{c^\theta d^2}\right)\int_0^t \frac{e^{\varphi(s)}}{(s+c)^{3b}} ds < \frac{e^{\varphi(t)}}{(c+t)^b}, \qquad \forall t > 0, \quad \theta = 1 - 2b > 0. \tag{18}$$

Proof. We have

$$\varphi(t) = \int_0^t \frac{d^2}{2(c+s)^{2b}} ds = \frac{d^2}{2(1-2b)}\Big((c+t)^{1-2b} - c^{1-2b}\Big) = p(c+t)^\theta - C_3, \tag{19}$$

where $\theta := 1 - 2b$, $p := \frac{d^2}{2\theta}$, $C_3 := pc^\theta$. One has

$$\begin{aligned}
\frac{d}{dt}\frac{e^{p(c+t)^\theta}}{(c+t)^b} &= \frac{p\theta e^{p(c+t)^\theta}}{(c+t)^{b+1-\theta}} - \frac{be^{p(c+t)^\theta}}{(c+t)^{b+1}} \\
&= \frac{e^{p(c+t)^\theta}}{(c+t)^b}\left(\frac{d^2}{2(c+t)^{2b}} - \frac{b}{c+t}\right) \\
&\geq \frac{e^{p(c+t)^\theta}}{(c+t)^b}\frac{d^2}{2(c+t)^{2b}}\left(1 - \frac{2b}{c^\theta d^2}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{d^2}{2}\left(1 - \frac{2b}{c^\theta d^2}\right)\int_0^t \frac{e^{p(c+s)^\theta}}{(s+c)^{3b}} ds &\leq \int_0^t \frac{d}{ds}\frac{e^{p(c+s)^\theta}}{(c+s)^b} ds \\
&\leq \frac{e^{p(c+t)^\theta}}{(c+t)^b} - \frac{e^{pc^\theta}}{c^b} \leq \frac{e^{p(c+t)^\theta}}{(c+t)^b}.
\end{aligned}$$

Multiplying this inequality by e^{-C_3} and using (19), one obtains (18). Lemma 8 is proved. $\Box$

Lemma 9 *Let* $a(t) = \frac{d}{(c+t)^b}$ *and* $\varphi(t) := \int_0^t \frac{a^2(s)}{2} ds$ *where* $d, c > 0$, $b \in (0, \frac{1}{4}]$ *and* $c^{1-2b}d^2 \geq 6b$. *One has*

$$e^{-\varphi(t)}\int_0^t e^{\varphi(s)}|\dot{a}(s)|\|V_\delta(s)\| ds \leq \frac{1}{2}a(t)\|V_\delta(t)\|, \qquad t \geq 0. \tag{20}$$

Proof. From Lemma 8, one has

$$\frac{1}{2}\left(1 - \frac{2b}{c^\theta d^2}\right)\int_0^t e^{\varphi(s)}\frac{d^3}{(s+c)^{3b}} ds < e^{\varphi(t)}\frac{d}{(c+t)^b}, \qquad \forall c, b \geq 0, \quad \theta = 1 - 2b > 0. \tag{21}$$

Since $c^{1-2b}d^2 \geq 6b$ or $\frac{6b}{c^\theta c_1^2} \leq 1$, one has

$$1 - \frac{2b}{c^\theta d^2} \geq \frac{4b}{c^\theta d^2} \geq \frac{4b}{(c+s)^{1-2b}d^2}, \qquad s \geq 0.$$

This implies

$$\frac{a^3(s)}{2}\left(1-\frac{2b}{c^\theta d^2}\right)=\frac{d^3}{2(c+s)^{3b}}\left(1-\frac{2b}{c^\theta d^2}\right)\geq\frac{4db}{2(c+s)^{b+1}}=2|\dot a(s)|,\qquad s\geq 0. \tag{22}$$

Multiplying (21) by $\|V_\delta(t)\|$, using inequality (22) and the fact that $\|V_\delta(t)\|$ is increasing, one gets, for all $t>0$, the following inequalities:

$$e^{\varphi(t)}a(t)\|V_\delta(t)\|>\int_0^t e^{\varphi(s)}\|V_\delta(t)\|\frac{a^3(s)}{2}\left(1-\frac{2b}{c^\theta d^2}\right)ds\geq 2\int_0^t e^{\varphi(s)}|\dot a(s)|\|V_\delta(s)\|ds.$$

This implies inequality (20). Lemma 9 is proved. □

Let us recall the following lemma, which is basic in our proofs.

Lemma 10 ([10], p. 97) *Let $\alpha(t)$, $\beta(t)$, $\gamma(t)$ be continuous nonnegative functions on $[t_0,\infty)$, $t_0\geq 0$ is a fixed number. If there exists a function*

$$\mu\in C^1[t_0,\infty),\quad \mu>0,\quad \lim_{t\to\infty}\mu(t)=\infty,$$

such that

$$0\leq\alpha(t)\leq\frac{\mu}{2}\left[\gamma-\frac{\dot\mu(t)}{\mu(t)}\right],\qquad \dot\mu:=\frac{d\mu}{dt}, \tag{23}$$

$$\beta(t)\leq\frac{1}{2\mu}\left[\gamma-\frac{\dot\mu(t)}{\mu(t)}\right], \tag{24}$$

$$\mu(0)g(0)<1, \tag{25}$$

and $g(t)\geq 0$ satisfies the inequality

$$\dot g(t)\leq-\gamma(t)g(t)+\alpha(t)g^2(t)+\beta(t),\quad t\geq t_0, \tag{26}$$

then $g(t)$ exists on $[t_0,\infty)$ and

$$0\leq g(t)<\frac{1}{\mu(t)}\to 0,\quad as\quad t\to\infty. \tag{27}$$

If inequalities (23)–(25) hold on an interval $[t_0,T)$, then $g(t)$ exists on this interval and inequality (27) holds on $[t_0,T)$.

Lemma 11 *Suppose M_1, c_0, and c_1 are positive constants and $0\neq y\in H$. Then there exist $\lambda>0$ and a function $a(t)\in C^1[0,\infty)$, $0<a(t)\searrow 0$, such that*

$$|\dot a(t)|\leq\frac{a^3(t)}{4},$$

and the following conditions hold

$$\frac{M_1}{\|y\|} \le \lambda, \tag{28}$$

$$c_0(M_1 + a(t)) \le \frac{\lambda}{2a^2(t)}\left[a^2(t) - \frac{2|\dot{a}(t)|}{a(t)}\right], \tag{29}$$

$$c_1\frac{|\dot{a}(t)|}{a(t)} \le \frac{a^2(t)}{2\lambda}\left[a^2(t) - \frac{2|\dot{a}(t)|}{a(t)}\right], \tag{30}$$

$$\frac{\lambda}{a^2(0)}g(0) < 1. \tag{31}$$

Proof. Take

$$a(t) = \frac{d}{(c+t)^b}, \quad 0 < b \le \frac{1}{4}, \quad 4b \le c^{1-2b}d^2, \quad c \ge 1. \tag{32}$$

Note that $|\dot{a}| = -\dot{a}$. We have

$$\frac{|\dot{a}|}{a^3} = \frac{b}{d^2(c+t)^{1-2b}} \le \frac{b}{d^2c^{1-2b}} \le \frac{1}{4}.$$

Hence,

$$\frac{a^2(t)}{2} \le a^2(t) - \frac{2|\dot{a}(t)|}{a(t)}. \tag{33}$$

Thus, inequality (29) is satisfied if

$$c_0(M_1 + a(0)) \le \frac{\lambda}{4}. \tag{34}$$

Take

$$\lambda \ge \max\left(8c_0M_1, \frac{M_1}{\|y\|}\right). \tag{35}$$

Then (28) is satisfied and

$$c_0M_1 \le \frac{\lambda}{8}. \tag{36}$$

For any given $g(0)$, choose $a(0)$ sufficiently large so that

$$\frac{\lambda}{a^2(0)}g(0) < 1.$$

Then inequality (31) is satisfied.

Choose $\kappa \ge 1$ such that

$$\kappa > \max\left(\sqrt{\frac{4\lambda c_1 b}{d^4}}, \frac{8c_0a(0)}{\lambda}, 1\right). \tag{37}$$

Define

$$\nu(t) := \kappa a(t), \qquad \lambda_\kappa := \kappa^2\lambda. \tag{38}$$

Using inequalities (36), (37) and (38), one gets

$$c_0(M_1 + \nu(0)) \le \frac{\lambda}{8} + c_0\nu(0) \le \frac{\lambda_\kappa}{8} + \frac{\lambda_\kappa}{8} = \frac{\lambda_\kappa}{4}.$$

Thus, (34) holds for $a(t) = \nu(t)$, $\lambda = \lambda_\kappa$. Consequently, (29) holds for $a(t) = \nu(t)$, $\lambda = \lambda_\kappa$ since (33) holds as well under this transformation, i.e.,

$$\frac{\nu^2(t)}{2} \le \nu^2(t) - \frac{2|\dot\nu(t)|}{\nu(t)}. \tag{39}$$

Using the inequalities (37) and $c \ge 1$ and the definition (38), one obtains

$$4\lambda_\kappa c_1 \frac{|\dot\nu(t)|}{\nu^5(t)} = 4\lambda c_1 \frac{b}{\kappa^2 d^4 (c+t)^{1-4b}} \le 4\lambda c_1 \frac{b}{\kappa^2 d^4} \le 1.$$

This implies

$$c_1 \frac{|\dot\nu|}{\nu(t)} \le \frac{\nu^4(t)}{4\lambda_\kappa} \le \frac{\nu^2(t)}{2\lambda_\kappa}\left[\nu^2 - \frac{2|\dot\nu|}{\nu}\right].$$

Thus, one can replace the function $a(t)$ by $\nu(t) = \kappa a(t)$ and λ by $\lambda_\kappa = \kappa^2\lambda$ in the inequalities (28)–(31). □

Lemma 12 *Suppose M_1, c_0, c_1 and $\tilde\alpha$ are positive constants and $0 \ne y \in H$. Then there exist $\lambda > 0$ and a sequence $0 < (a_n)_{n=0}^\infty \searrow 0$ such that the following conditions hold*

$$\frac{a_n}{a_{n+1}} \le 2, \tag{40}$$

$$\|f_\delta - F(0)\| \le \frac{a_0^3}{\lambda}, \tag{41}$$

$$\frac{M_1}{\lambda} \le \|y\|, \tag{42}$$

$$\frac{c_0(M_1 + a_0)}{\lambda} \le \frac{1}{2}, \tag{43}$$

$$\frac{a_n^2}{\lambda} - \frac{\tilde\alpha a_n^4}{2\lambda} + \frac{a_n - a_{n+1}}{a_{n+1}} c_1 \le \frac{a_{n+1}^2}{\lambda}. \tag{44}$$

Proof. Let us show that if $a_0 > 0$ is sufficiently large, then the following sequence

$$a_n = \frac{a_0}{(1+n)^b}, \qquad b = \frac{1}{4}, \tag{45}$$

satisfies conditions (41)–(44) if

$$\lambda \ge \max\left(\frac{M_1}{\|y\|}, 4c_0 M_1\right). \tag{46}$$

Condition (40) is satisfied by the sequence (45). Inequality (42) is satisfied since (46) holds. Choose $a(0)$ so that

$$a_0 \ge \sqrt[3]{\|f_\delta - F(0)\|\lambda}, \tag{47}$$

then (41) is satisfied.

Assume that $(a_n)_{n=0}^{\infty}$ and λ satisfy (40), (41) and (42). Choose $\kappa \geq 1$ such that

$$\kappa \geq \max\left(\frac{4c_0a_0}{\lambda}, \sqrt{\frac{4}{\tilde{\alpha}a_0^2 2\sqrt{2}}}, \sqrt{\frac{\lambda c_1}{\tilde{\alpha}a_0^4}}\right). \tag{48}$$

It follows from (48) that

$$\frac{4}{\kappa^2 a_0^2 2\sqrt{2}} \leq \tilde{\alpha}, \qquad \frac{\lambda c_1}{\kappa^2 a_0^4} \leq \tilde{\alpha}. \tag{49}$$

Define

$$(b_n)_{n=0}^{\infty} := (\kappa a_n)_{n=0}^{\infty}, \qquad \lambda_\kappa := \kappa^2\lambda. \tag{50}$$

Using inequalities (46), (48) and the definitions (50), one gets

$$\frac{c_0(M_1+b_0)}{\lambda_\kappa} \leq \frac{1}{4} + \frac{c_0a_0}{\kappa\lambda} \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Thus, inequality (43) holds for a_0 replaced by $b_0 = \kappa a_0$ and λ replaced by $\lambda_\kappa = \kappa^2\lambda$, where κ satisfies (48).

For all $n \geq 0$ one has

$$\frac{a_n^2 - a_{n+1}^2}{a_n^4} = \frac{a_n^4 - a_{n+1}^4}{a_n^4(a_n^2 + a_{n+1}^2)} \leq \frac{a_n^4 - a_{n+1}^4}{2a_{n+1}^2 a_n^4} = \frac{\frac{a_0^4}{n+1} - \frac{a_0^4}{n+2}}{2\frac{a_0^2}{\sqrt{n+2}}\frac{a_0^4}{n+1}} = \frac{1}{a_0^2 2\sqrt{n+2}} \leq \frac{1}{a_0^2 2\sqrt{2}}. \tag{51}$$

Since a_n is decreasing, one has

$$\begin{aligned} \frac{a_n - a_{n+1}}{a_n^4 a_{n+1}} &= \frac{a_n^4 - a_{n+1}^4}{a_n^4 a_{n+1}(a_n + a_{n+1})(a_n^2 + a_{n+1}^2)} \\ &\leq \frac{a_n^4 - a_{n+1}^4}{4a_n^4 a_{n+1}^4} = \frac{\frac{a_0^4}{n+1} - \frac{a_0^4}{n+2}}{4\frac{a_0^4}{n+2}\frac{a_0^4}{n+1}} \leq \frac{1}{4a_0^4}, \qquad \forall n \geq 0. \end{aligned} \tag{52}$$

Using inequalities (51) and (49), one gets

$$\frac{4(a_n^2 - a_{n+1}^2)}{\kappa^2 a_n^4} \leq \frac{4}{\kappa^2 a_0^2 2\sqrt{2}} \leq \tilde{\alpha}. \tag{53}$$

Similarly, using inequalities (52) and (49), one gets

$$\frac{4\lambda(a_n - a_{n+1})c_1}{\kappa^2 a_n^4 a_{n+1}} \leq \frac{\lambda c_1}{\kappa^2 a_0^4} \leq \tilde{\alpha}. \tag{54}$$

Inequalities (53) and (54) imply

$$\begin{aligned} \frac{b_n^2 - b_{n+1}^2}{\lambda_\kappa} + \frac{b_n - b_{n+1}}{b_{n+1}}c_1 &= \frac{a_n^2 - a_{n+1}^2}{\lambda} + \frac{a_n - a_{n+1}}{a_{n+1}}c_1 \\ &= \frac{\kappa^2 a_n^4}{4\lambda}\frac{4(a_n^2 - a_{n+1}^2)}{\kappa^2 a_n^4} + \frac{\kappa^2 a_n^4}{4\lambda}\frac{4\lambda(a_n - a_{n+1})c_1}{\kappa^2 a_n^4 a_{n+1}} \\ &\leq \frac{\kappa^2 a_n^4}{4\lambda}\tilde{\alpha} + \frac{\kappa^2 a_n^4}{4\lambda}\tilde{\alpha} = \frac{\kappa^2 a_n^4 \tilde{\alpha}}{2\lambda} = \frac{\tilde{\alpha} b_n^4}{2\lambda_\kappa}. \end{aligned}$$

Thus, inequality (44) holds for a_n replaced by $b_n = \kappa a_n$ and λ replaced by $\lambda_\kappa = \kappa^2\lambda$, where κ satisfies (48). Inequalities (40)–(42) hold as well under this transformation. Thus, the choices $a_n = b_n$ and $\lambda := \kappa \max\left(\frac{M_1}{\|y\|}, 4c_0 M_1\right)$, where κ satisfies (48), satisfy all the conditions of Lemma 12. □

Remark 13 The constant c_0 and c_1 used in Lemma 11 and 12 will be used in Theorems 17 and 19. These constants are defined in equation (67). The constant $\tilde{\alpha}$, used in Lemma 12, is the one from Theorem 19. This constant is defined in (94).

Remark 14 Using similar arguments one can show that the sequence $a_n = \frac{d}{(c+n)^b}$, where $c \geq 1$, $0 < b \leq \frac{1}{4}$, satisfy all conditions of Lemma 4 provided that d is sufficiently large and λ is chosen so that inequality (46) holds.

Remark 15 In the proof of Lemma 12 and 11 the numbers a_0 and λ can be chosen so that $\frac{a_0^2}{\lambda}$ is uniformly bounded as $\delta \to 0$ regardless of the rate of growth of the constant $M_1 = M_1(R)$ from formula (3) when $R \to \infty$, i.e., regardless of the strength of the nonlinearity $F(u)$.

To satisfy (46) one can choose $\lambda = M_1\big(\frac{1}{\|y\|} + 4c_0\big)$. To satisfy (47) one can choose

$$a_0 = \sqrt[3]{\lambda(\|f - F(0)\| + \|f\|)} \geq \sqrt[3]{\lambda\|f_\delta - F(0)\|},$$

where we have assumed without loss of generality that $0 < \|f_\delta - f\| < \|f\|$. With this choice of a_0 and λ, the ratio $\frac{a_0^2}{\lambda}$ is bounded uniformly with respect to $\delta \in (0,1)$ and does not depend on R. The dependence of a_0 on δ is seen from (47) since f_δ depends on δ. In practice one has $\|f_\delta - f\| < \|f\|$. Consequently,

$$\sqrt[3]{\|f_\delta - F(0)\|\lambda} \leq \sqrt[3]{(\|f - F(0)\| + \|f\|)\lambda}.$$

Thus, we can practically choose $a(0)$ independent of δ from the following inequality

$$a_0 \geq \sqrt[3]{\lambda(\|f - F(0)\| + \|f\|)}.$$

Indeed, with the above choice one has $\frac{a_0^2}{\lambda} \leq c(1 + \sqrt[3]{\lambda^{-1}}) \leq c$, where $c > 0$ is a constant independent of δ, and one can assume that $\lambda \geq 1$ without loss of generality.

This Remark is used in the proof of the main result in Section 3. Specifically, it is used to prove that an iterative process (93) generates a sequence which stays in the ball $B(u_0, R)$ for all $n \leq n_0 + 1$, where the number n_0 is defined by formula (104) (see below), and $R > 0$ is sufficiently large. An upper bound on R is given in the proof of Theorem 19, below formula (117).

Remark 16 One can choose $u_0 \in H$ such that

$$g_0 := \|u_0 - V_0\| \leq \frac{\|F(0) - f_\delta\|}{a_0}. \tag{55}$$

Indeed, if, for example, $u_0 = 0$, then by Remark 6 one gets

$$g_0 = \|V_0\| = \frac{a_0\|V_0\|}{a_0} \le \frac{\|F(0) - f_\delta\|}{a_0}.$$

If (41) and (55) hold then $g_0 \le \frac{a_0^2}{\lambda}$.

3 Main results

3.1 Dynamical systems gradient method

Assume:

$$0 < a(t) \searrow 0, \quad \lim_{t\to\infty} \frac{\dot a(t)}{a(t)} = 0, \quad \frac{|\dot a(t)|}{a^3(t)} \le \frac{1}{4}. \tag{56}$$

Denote

$$A := F'(u_\delta(t)), \quad A_a := A + aI, \quad a = a(t),$$

where I is the identity operator, and $u_\delta(t)$ solves the following Cauchy problem:

$$\dot u_\delta = -A^*_{a(t)}[F(u_\delta) + a(t)u_\delta - f_\delta], \quad u_\delta(0) = u_0. \tag{57}$$

Theorem 17 *Assume that $F : H \to H$ is a monotone operator, twice Fréchet differentiable, $\sup_{u\in B(u_0,R)} \|F^{(j)}(u)\| \le M_j(R)$, $0 \le j \le 2$, $B(u_0, R) := \{u : \|u - u_0\| \le R\}$, u_0 is an element of H, satisfying inequality (88) (see below). Let $a(t)$ satisfy conditions of Lemma 11. For example, one can choose $a(t) = \frac{d}{(c+t)^b}$, where $b \in (0, \frac{1}{4}]$, $c \ge 1$, and $d > 0$ are constants, and d is sufficiently large. Assume that equation $F(u) = f$ has a solution in $B(u_0, R)$, possibly nonunique, and y is the minimal-norm solution to this equation. Let f be unknown but f_δ be given, $\|f_\delta - f\| \le \delta$. Then the solution $u_\delta(t)$ to problem (57) exists on an interval $[0, T_\delta]$, $\lim_{\delta\to 0} T_\delta = \infty$, and there exists t_δ, $t_\delta \in (0, T_\delta)$, not necessarily unique, such that*

$$\|F(u_\delta(t_\delta)) - f_\delta\| = C_1\delta^\zeta, \quad \lim_{\delta\to 0} t_\delta = \infty, \tag{58}$$

where $C_1 > 1$ and $0 < \zeta \le 1$ are constants. If $\zeta \in (0, 1)$ and t_δ satisfies (58), then

$$\lim_{\delta\to 0} \|u_\delta(t_\delta) - y\| = 0. \tag{59}$$

Remark 18 One can easily choose u_0 satisfying inequality (88). Note that inequality (88) is a sufficient condition for (91) to hold. In our proof inequality (91) is used at $t = t_\delta$. The stopping time t_δ is often sufficiently large for the quantity $e^{-\varphi(t_\delta)}h_0$ to be small. In this case inequality (91) with $t = t_\delta$ is satisfied for a wide range of u_0. The parameter ζ is not fixed in (58). While we could fix it, for example, by setting $\zeta = 0.9$, it is an interesting *open problem* to propose an optimal in some sense criterion for choosing ζ.

Proof. [Proof of Theorem 17] Denote

$$C := \frac{C_1 + 1}{2}. \tag{60}$$

Let

$$w := u_\delta - V_\delta, \quad g(t) := \|w\|.$$

One has

$$\dot{w} = -\dot{V}_\delta - A^*_{a(t)}\big[F(u_\delta) - F(V_\delta) + a(t)w\big]. \tag{61}$$

We use Taylor's formula and get:

$$F(u_\delta) - F(V_\delta) + aw = A_a w + K, \quad \|K\| \le \frac{M_2}{2}\|w\|^2, \tag{62}$$

where $K := F(u_\delta) - F(V_\delta) - Aw$, and M_2 is the constant from the estimate (3). Multiplying (61) by w and using (62) one gets

$$g\dot{g} \le -a^2 g^2 + \frac{M_2(M_1+a)}{2} g^3 + \|\dot{V}_\delta\| g, \tag{63}$$

where the estimates: $\langle A^*_a A_a w, w\rangle \ge a^2 g^2$ and $\|A_a\| \le M_1 + a$ were used. Note that the inequality $\langle A^*_a A_a w, w\rangle \ge a^2 g^2$ is true if $A \ge 0$. Since F is monotone and differentiable (see (3)), one has $A := F'(u_\delta) \ge 0$.

Let $t_0 > 0$ be such that

$$\frac{\delta}{a(t_0)} = \frac{1}{C-1}\|y\|, \qquad C > 1. \tag{64}$$

This t_0 exists and is unique since $a(t) > 0$ monotonically decays to 0 as $t \to \infty$. By Lemma 4, there exists t_1 such that

$$\|F(V_\delta(t_1)) - f_\delta\| = C\delta, \quad F(V_\delta(t_1)) + a(t_1)V_\delta(t_1) - f_\delta = 0. \tag{65}$$

We claim that $t_1 \in [0, t_0]$.

Indeed, from (8) and (17) one gets

$$C\delta = a(t_1)\|V_\delta(t_1)\| \le a(t_1)\bigg(\|y\| + \frac{\delta}{a(t_1)}\bigg) = a(t_1)\|y\| + \delta, \quad C > 1,$$

so

$$\delta \le \frac{a(t_1)\|y\|}{C-1}.$$

Thus,

$$\frac{\delta}{a(t_1)} \le \frac{\|y\|}{C-1} = \frac{\delta}{a(t_0)}.$$

Since $a(t) \searrow 0$, the above inequality implies $t_1 \le t_0$. Differentiating both sides of (8) with respect to t, one obtains

$$A_{a(t)}\dot{V}_\delta = -\dot{a}V_\delta.$$

This implies

$$\|\dot{V}_\delta\| \le |\dot{a}|\|A^{-1}_{a(t)}V_\delta\| \le \frac{|\dot{a}|}{a}\|V_\delta\| \le \frac{|\dot{a}|}{a}\bigg(\|y\| + \frac{\delta}{a}\bigg) \le \frac{|\dot{a}|}{a}\|y\|\bigg(1 + \frac{1}{C-1}\bigg), \quad \forall t \le t_0. \tag{66}$$

Since $g \geq 0$, inequalities (63) and (66) imply

$$\dot{g} \leq -a^2(t)g(t) + c_0(M_1 + a(t))g^2 + \frac{|\dot{a}(t)|}{a(t)}c_1, \quad c_0 = \frac{M_2}{2},\, c_1 = \|y\|\left(1 + \frac{1}{C-1}\right). \tag{67}$$

Inequality (67) is of the type (26) with

$$\gamma(t) = a^2(t), \quad \alpha(t) = c_0(M_1 + a(t)), \quad \beta(t) = c_1\frac{|\dot{a}(t)|}{a(t)}.$$

Let us check assumptions (23)–(25). Take

$$\mu(t) = \frac{\lambda}{a^2(t)}, \quad \lambda = \text{const}.$$

By Lemma 11 there exist λ and $a(t)$ such that conditions (23)–(25) hold. Thus, Lemma 10 yields

$$g(t) < \frac{a^2(t)}{\lambda}, \quad \forall t \leq t_0. \tag{68}$$

Therefore,

$$\begin{aligned} \|F(u_\delta(t)) - f_\delta\| \leq & \|F(u_\delta(t)) - F(V_\delta(t))\| + \|F(V_\delta(t)) - f_\delta\| \\ \leq & M_1 g(t) + \|F(V_\delta(t)) - f_\delta\| \\ \leq & \frac{M_1 a^2(t)}{\lambda} + \|F(V_\delta(t)) - f_\delta\|, \qquad \forall t \leq t_0. \end{aligned} \tag{69}$$

It follows from Lemma 3 that $\|F(V_\delta(t)) - f_\delta\|$ is decreasing. Since $t_1 \leq t_0$, one gets

$$\|F(V_\delta(t_0)) - f_\delta\| \leq \|F(V_\delta(t_1)) - f_\delta\| = C\delta. \tag{70}$$

This, inequality (69), the inequality $\frac{M_1}{\lambda} \leq \|y\|$ (see (35)), the relation (64), and the definition $C_1 = 2C - 1$ (see (60)) imply

$$\begin{aligned} \|F(u_\delta(t_0)) - f_\delta\| \leq & \frac{M_1 a^2(t_0)}{\lambda} + C\delta \\ \leq & \frac{M_1 \delta(C-1)}{\lambda\|y\|} + C\delta \leq (2C-1)\delta = C_1\delta. \end{aligned} \tag{71}$$

We have used the inequality

$$a^2(t_0) \leq a(t_0) = \frac{\delta(C-1)}{\|y\|}$$

which is true if δ is sufficiently small, or, equivalently, if t_0 is sufficiently large. Thus, if

$$\|F(u_\delta(0)) - f_\delta\| \geq C_1\delta^\zeta, \quad 0 < \zeta \leq 1,$$

then there exists $t_\delta \in (0, t_0)$ such that

$$\|F(u_\delta(t_\delta)) - f_\delta\| = C_1\delta^\zeta \tag{72}$$

for any given $\zeta \in (0,1]$, and any fixed $C_1 > 1$.

Let us prove (59). *If this is done, then Theorem 17 is proved.*

First, we prove that $\lim_{\delta\to 0} \frac{\delta}{a(t_\delta)} = 0$.

From (69) with $t = t_\delta$, and from (17), one gets

$$\begin{aligned} C_1\delta^\zeta &\le M_1\frac{a^2(t_\delta)}{\lambda} + a(t_\delta)\|V_\delta(t_\delta)\| \\ &\le M_1\frac{a^2(t_\delta)}{\lambda} + \|y\|a(t_\delta) + \delta. \end{aligned}$$

Thus, for sufficiently small δ, one gets

$$\tilde{C}\delta^\zeta \le a(t_\delta)\Big(\frac{M_1 a(0)}{\lambda} + \|y\|\Big), \quad \tilde{C} > 0,$$

where $\tilde{C} < C_1$ is a constant. Therefore,

$$\lim_{\delta\to 0}\frac{\delta}{a(t_\delta)} \le \lim_{\delta\to 0}\frac{\delta^{1-\zeta}}{\tilde{C}}\Big(\frac{M_1 a(0)}{\lambda} + \|y\|\Big) = 0, \quad 0 < \zeta < 1. \tag{73}$$

Secondly, we prove that

$$\lim_{\delta\to 0} t_\delta = \infty. \tag{74}$$

Using (57), one obtains:

$$\frac{d}{dt}\big(F(u_\delta) + au_\delta - f_\delta\big) = A_a\dot{u}_\delta + \dot{a}u_\delta = -A_aA_a^*\big(F(u_\delta) + au_\delta - f_\delta\big) + \dot{a}u_\delta.$$

This and (8) imply:

$$\frac{d}{dt}\big[F(u_\delta) - F(V_\delta) + a(u_\delta - V_\delta)\big] = -A_aA_a^*\big[F(u_\delta) - F(V_\delta) + a(u_\delta - V_\delta)\big] + \dot{a}u_\delta. \tag{75}$$

Denote

$$v := F(u_\delta) - F(V_\delta) + a(u_\delta - V_\delta), \quad h = \|v\|.$$

Multiplying (75) by v and using monotonicity of F, one obtains

$$\begin{aligned} h\dot{h} &= -\langle A_aA_a^*v, v\rangle + \langle v, \dot{a}(u_\delta - V_\delta)\rangle + \dot{a}\langle v, V_\delta\rangle \\ &\le -h^2a^2 + h|\dot{a}|\|u_\delta - V_\delta\| + |\dot{a}|h\|V_\delta\|, \qquad h \ge 0. \end{aligned} \tag{76}$$

Again, we have used the inequality $A_aA_a^* \ge a^2$, which holds for $A \ge 0$, i.e., monotone operators F. Thus,

$$\dot{h} \le -ha^2 + |\dot{a}|\|u_\delta - V_\delta\| + |\dot{a}|\|V_\delta\|. \tag{77}$$

Since $\langle F(u_\delta) - F(V_\delta), u_\delta - V_\delta\rangle \ge 0$, one obtains two inequalities

$$a\|u_\delta - V_\delta\|^2 \le \langle v, u_\delta - V_\delta\rangle \le \|u_\delta - V_\delta\|h, \tag{78}$$

and

$$\|F(u_\delta) - F(V_\delta)\|^2 \le \langle v, F(u_\delta) - F(V_\delta)\rangle \le h\|F(u_\delta) - F(V_\delta)\|. \tag{79}$$

Inequalities (78) and (79) imply:

$$a\|u_\delta - V_\delta\| \le h, \quad \|F(u_\delta) - F(V_\delta)\| \le h. \tag{80}$$

Inequalities (77) and (80) imply

$$\dot{h} \le -h\Big(a^2 - \frac{|\dot{a}|}{a}\Big) + |\dot{a}|\|V_\delta\|. \tag{81}$$

Since $a^2 - \frac{|\dot{a}|}{a} \ge \frac{3a^2}{4} > \frac{a^2}{2}$ by the last inequality in (56), it follows from inequality (81) that

$$\dot{h} \le -\frac{a^2}{2}h + |\dot{a}|\|V_\delta\|. \tag{82}$$

Inequality (82) implies:

$$h(t) \le h(0)e^{-\int_0^t \frac{a^2(s)}{2}ds} + e^{-\int_0^t \frac{a^2(s)}{2}ds}\int_0^t e^{\int_0^s \frac{a^2(\xi)}{2}d\xi}|\dot{a}(s)|\|V_\delta(s)\|ds. \tag{83}$$

Denote

$$\varphi(t) := \int_0^t \frac{a^2(s)}{2}ds.$$

From (83) and (80), one gets

$$\|F(u_\delta(t)) - F(V_\delta(t))\| \le h(0)e^{-\varphi(t)} + e^{-\varphi(t)}\int_0^t e^{\varphi(s)}|\dot{a}(s)|\|V_\delta(s)\|ds. \tag{84}$$

Therefore,

$$\begin{aligned}\|F(u_\delta(t)) - f_\delta\| &\ge \|F(V_\delta(t)) - f_\delta\| - \|F(V_\delta(t)) - F(u_\delta(t))\| \\ &\ge a(t)\|V_\delta(t)\| - h(0)e^{-\varphi(t)} - e^{-\varphi(t)}\int_0^t e^{\varphi(s)}|\dot{a}|\|V_\delta\|ds.\end{aligned} \tag{85}$$

From Lemma 9 it follows that there exists an $a(t)$ such that

$$\frac{1}{2}a(t)\|V_\delta(t)\| \ge e^{-\varphi(t)}\int_0^t e^{\varphi(s)}|\dot{a}|\|V_\delta(s)\|ds. \tag{86}$$

For example, one can choose

$$a(t) = \frac{c_1}{(c+t)^b}, \quad b \in (0, \frac{1}{4}], \quad c_1^2 c^{1-2b} \ge 6b, \tag{87}$$

where $c_1, c > 0$. Moreover, one can always choose u_0 such that

$$h(0) = \|F(u_0) + a(0)u_0 - f_\delta\| \le \frac{1}{4}a(0)\|V_\delta(0)\|, \tag{88}$$

because the equation

$$F(u_0) + a(0)u_0 - f_\delta = 0$$

is solvable.

If (88) holds, then

$$h(0)e^{-\varphi(t)} \le \frac{1}{4}a(0)\|V_\delta(0)\|e^{-\varphi(t)}, \qquad t \ge 0. \tag{89}$$

If (87) holds, $c \ge 1$ and $2b \le c_1^2$, then it follows that

$$e^{-\varphi(t)}a(0) \le a(t). \tag{90}$$

Indeed, inequality $a(0) \le a(t)e^{\varphi(t)}$ is obviously true for $t = 0$, and $\big(a(t)e^{\varphi(t)}\big)'_t \ge 0$, provided that $c \ge 1$ and $2b \le c_1^2$.

Inequalities (89) and (50) imply

$$e^{-\varphi(t)}h(0) \le \frac{1}{4}a(t)\|V_\delta(0)\| \le \frac{1}{4}a(t)\|V_\delta(t)\|, \quad t \ge 0. \tag{91}$$

where we have used the inequality $\|V_\delta(t)\| \le \|V_\delta(t')\|$ for $t \le t'$, established in Lemma 3. From (72) and (85)–(91), one gets

$$C\delta^\zeta = \|F(u_\delta(t_\delta)) - f_\delta\| \ge \frac{1}{4}a(t_\delta)\|V_\delta(t_\delta)\|.$$

Thus,

$$\lim_{\delta\to 0} a(t_\delta)\|V_\delta(t_\delta)\| \le \lim_{\delta\to 0} 4C\delta^\zeta = 0.$$

Since $\|V_\delta(t)\|$ is increasing, this implies $\lim_{\delta\to 0} a(t_\delta) = 0$. Since $0 < a(t) \searrow 0$, it follows that (74) holds.

From the triangle inequality and inequalities (68) and (15) one obtains

$$\begin{aligned}\|u_\delta(t_\delta) - y\| &\le \|u_\delta(t_\delta) - V_\delta\| + \|V(t_\delta) - V_\delta(t_\delta)\| + \|V(t_\delta) - y\| \\ &\le \frac{a^2(t_\delta)}{\lambda} + \frac{\delta}{a(t_\delta)} + \|V(t_\delta) - y\|.\end{aligned} \tag{92}$$

From (73), (74), inequality (92) and Lemma 1, one obtains (59). Theorem 17 is proved. □

3.2 An iterative scheme

Let $V_{n,\delta}$ solve the equation:

$$F(V_{n,\delta}) + a_n V_{n,\delta} - f_\delta = 0.$$

Denote $V_n := V_{n,\delta}$.

Consider the following iterative scheme:

$$u_{n+1} = u_n - \alpha_n A_n^*[F(u_n) + a_n u_n - f_\delta], \quad A_n := F'(u_n) + a_n I, \quad u_0 = u_0, \tag{93}$$

where u_0 is chosen so that inequality (55) holds, and $\{\alpha_n\}_{n=1}^{\infty}$ is a positive sequence such that

$$0 < \tilde{\alpha} \le \alpha_n \le \frac{2}{a_n^2 + (M_1 + a_n)^2}, \qquad ||A_n|| \le M_1 + a_n. \tag{94}$$

It follows from this condition that

$$\|1 - \alpha_n A_{a_n}^* A_{a_n}\| = \sup_{a_n^2 \le \lambda \le (M_1 + a_n)^2} |1 - \alpha_n \lambda| \le 1 - \alpha_n a_n^2. \tag{95}$$

Note that $F'(u_n) \ge 0$ since F is monotone.

Let a_n and λ satisfy conditions (40)–(44). Assume that equation $F(u) = f$ has a solution in $B(u_0, R)$, possibly nonunique, and y is the minimal-norm solution to this equation. Let f be unknown but f_δ be given, and $\|f_\delta - f\| \le \delta$. We prove the following result:

Theorem 19 *Assume $a_n = \frac{d}{(c+n)^b}$ where $c \ge 1$, $0 < b \le \frac{1}{4}$, and d is sufficiently large so that conditions (40)–(44) hold. Let u_n be defined by (93). Assume that u_0 is chosen so that (55) holds. Then there exists a unique n_δ such that*

$$\|F(u_{n_\delta}) - f_\delta\| \le C_1 \delta^\zeta, \quad C_1 \delta^\zeta < \|F(u_n) - f_\delta\|, \quad \forall n < n_\delta, \tag{96}$$

where $C_1 > 1$, $0 < \zeta \le 1$.

Let $0 < (\delta_m)_{m=1}^{\infty}$ be a sequence such that $\delta_m \to 0$. If the sequence $\{n_m := n_{\delta_m}\}_{m=1}^{\infty}$ is bounded, and $\{n_{m_j}\}_{j=1}^{\infty}$ is a convergent subsequence, then

$$\lim_{j \to \infty} u_{n_{m_j}} = \tilde{u}, \tag{97}$$

where $\tilde{u}$ is a solution to the equation $F(u) = f$. If

$$\lim_{m \to \infty} n_m = \infty, \tag{98}$$

where $\zeta \in (0, 1)$, then

$$\lim_{m \to \infty} \|u_{n_m} - y\| = 0. \tag{99}$$

Proof. Denote

$$C := \frac{C_1 + 1}{2}. \tag{100}$$

Let

$$z_n := u_n - V_n, \quad g_n := \|z_n\|.$$

We use Taylor's formula and get:

$$F(u_n) - F(V_n) + a_n z_n = A_n z_n + K_n, \quad \|K_n\| \le \frac{M_2}{2} \|z_n\|^2, \tag{101}$$

where $K_n := F(u_n) - F(V_n) - F'(u_n) z_n$ and M_2 is the constant from (3). From (93) and (101) one obtains

$$z_{n+1} = z_n - \alpha_n A_n^* A_n z_n - \alpha_n A_n^* K(z_n) - (V_{n+1} - V_n). \tag{102}$$

From (102), (101), (95), and the estimate $\|A_n\| \le M_1 + a_n$, one gets

$$\begin{aligned} g_{n+1} &\le g_n \|1 - \alpha_n A_n^* A_n\| + \frac{\alpha_n M_2(M_1 + a_n)}{2} g_n^2 + \|V_{n+1} - V_n\| \\ &\le g_n(1 - \alpha_n a_n^2) + \frac{\alpha_n M_2(M_1 + a_n)}{2} g_n^2 + \|V_{n+1} - V_n\|. \end{aligned} \tag{103}$$

Since $0 < a_n \searrow 0$, for any fixed $\delta > 0$ there exists n_0 such that

$$\frac{\delta}{a_{n_0+1}} > \frac{1}{C-1}\|y\| \ge \frac{\delta}{a_{n_0}}, \qquad C > 1. \tag{104}$$

By (40), one has $\frac{a_n}{a_{n+1}} \le 2$, $\forall\, n \ge 0$. This and (104) imply

$$\frac{2}{C-1}\|y\| \ge \frac{2\delta}{a_{n_0}} > \frac{\delta}{a_{n_0+1}} > \frac{1}{C-1}\|y\| \ge \frac{\delta}{a_{n_0}}, \qquad C > 1. \tag{105}$$

Thus,

$$\frac{2}{C-1}\|y\| > \frac{\delta}{a_n}, \quad \forall n \le n_0 + 1. \tag{106}$$

The number n_0, satisfying (106), exists and is unique since $a_n > 0$ monotonically decays to 0 as $n \to \infty$. By Remark 5, there exists a number n_1 such that

$$\|F(V_{n_1+1}) - f_\delta\| \le C\delta < \|F(V_{n_1}) - f_\delta\|, \tag{107}$$

where V_n solves the equation $F(V_n) + a_n V_n - f_\delta = 0$.

We claim that $n_1 \in [0, n_0]$.

Indeed, one has $\|F(V_{n_1}) - f_\delta\| = a_{n_1}\|V_{n_1}\|$, and $\|V_{n_1}\| \le \|y\| + \frac{\delta}{a_{n_1}}$ (cf. (17)), so

$$C\delta < a_{n_1}\|V_{n_1}\| \le a_{n_1}\left(\|y\| + \frac{\delta}{a_{n_1}}\right) = a_{n_1}\|y\| + \delta, \quad C > 1. \tag{108}$$

Therefore,

$$\delta < \frac{a_{n_1}\|y\|}{C-1}. \tag{109}$$

Thus, by (105),

$$\frac{\delta}{a_{n_1}} < \frac{\|y\|}{C-1} < \frac{\delta}{a_{n_0+1}}. \tag{110}$$

Here the last inequality is a consequence of (105). Since a_n decreases monotonically, inequality (110) implies $n_1 \le n_0$. One has

$$\begin{aligned} a_{n+1}\|V_n - V_{n+1}\|^2 &= \langle (a_{n+1} - a_n)V_n - F(V_n) + F(V_{n+1}), V_n - V_{n+1}\rangle \\ &\le \langle (a_{n+1} - a_n)V_n, V_n - V_{n+1}\rangle \\ &\le (a_n - a_{n+1})\|V_n\|\|V_n - V_{n+1}\|. \end{aligned} \tag{111}$$

By (17), $\|V_n\| \le \|y\| + \frac{\delta}{a_n}$, and, by (106), $\frac{\delta}{a_n} \le \frac{2\|y\|}{C-1}$ for all $n \le n_0+1$. Therefore,

$$\|V_n\| \le \|y\| \left(1 + \frac{2}{C-1}\right), \qquad \forall n \le n_0 + 1, \tag{112}$$

and, by (111),

$$\|V_n - V_{n+1}\| \le \frac{a_n - a_{n+1}}{a_{n+1}} \|V_n\| \le \frac{a_n - a_{n+1}}{a_{n+1}} \|y\| \left(1 + \frac{2}{C-1}\right), \quad \forall n \le n_0 + 1. \tag{113}$$

Inequalities (103) and (113) imply

$$g_{n+1} \le (1 - \alpha_n a_n^2) g_n + \alpha_n c_0 (M_1 + a_n) g_n^2 + \frac{a_n - a_{n+1}}{a_{n+1}} c_1, \qquad \forall\, n \le n_0 + 1, \tag{114}$$

where the constants c_0 and c_1 are defined in (67).

By Lemma 4 and Remark 14, the sequence $(a_n)_{n=1}^{\infty}$, satisfies conditions (40)–(44), provided that a_0 is sufficiently large and $\lambda > 0$ is chosen so that (46) holds. Let us show by induction that

$$g_n < \frac{a_n^2}{\lambda}, \qquad 0 \le n \le n_0 + 1. \tag{115}$$

Inequality (115) holds for $n = 0$ by Remark 16. Suppose (115) holds for some $n \ge 0$. From (114), (115) and (44), one gets

$$\begin{aligned} g_{n+1} &\le (1 - \alpha_n a_n^2) \frac{a_n^2}{\lambda} + \alpha_n c_0 (M_1 + a_n) \left(\frac{a_n^2}{\lambda}\right)^2 + \frac{a_n - a_{n+1}}{a_{n+1}} c_1 \\ &= \frac{a_n^4}{\lambda} \left(\frac{\alpha_n c_0 (M_1 + a_n)}{\lambda} - \alpha_n\right) + \frac{a_n^2}{\lambda} + \frac{a_n - a_{n+1}}{a_{n+1}} c_1 \\ &\le -\frac{\alpha_n a_n^4}{2\lambda} + \frac{a_n^2}{\lambda} + \frac{a_n - a_{n+1}}{a_{n+1}} c_1 \\ &\le \frac{a_{n+1}^2}{\lambda}. \end{aligned} \tag{116}$$

Thus, by induction, inequality (115) holds for all n in the region $0 \le n \le n_0 + 1$.

From (17) one has $\|V_n\| \le \|y\| + \frac{\delta}{a_n}$. This and the triangle inequality imply

$$\|u_0 - u_n\| \le \|u_0\| + \|z_n\| + \|V_n\| \le \|u_0\| + \|z_n\| + \|y\| + \frac{\delta}{a_n}. \tag{117}$$

Inequalities (112), (115), and (117) guarantee that the sequence u_n, generated by the iterative process (93), remains in the ball $B(u_0, R)$ for all $n \le n_0 + 1$, where $R \le \frac{a_0}{\lambda} + \|u_0\| + \|y\| + \frac{\delta}{a_n}$. This inequality and the estimate (106) imply that the sequence u_n, $n \le n_0 + 1$, stays in the ball $B(u_0, R)$, where

$$R \le \frac{a_0}{\lambda} + \|u_0\| + \|y\| + \|y\| \frac{C+1}{C-1}. \tag{118}$$

By Remark 15, one can choose a_0 and λ so that $\frac{a_0}{\lambda}$ is uniformly bounded as $\delta \to 0$ even if $M_1(R) \to \infty$ as $R \to \infty$ at an arbitrary fast rate. Thus, the sequence u_n stays in the ball $B(u_0, R)$ for $n \le n_0 + 1$ when $\delta \to 0$. An upper bound on R is given above. It does not depend on δ as $\delta \to 0$.

One has:

$$\begin{aligned} \|F(u_n) - f_\delta\| \le& \|F(u_n) - F(V_n)\| + \|F(V_n) - f_\delta\| \\ \le& M_1 g_n + \|F(V_n) - f_\delta\| \\ \le& \frac{M_1 a_n^2}{\lambda} + \|F(V_n) - f_\delta\|, \qquad \forall n \le n_0 + 1, \end{aligned} \tag{119}$$

where (115) was used and M_1 is the constant from (3). Since $\|F(V_n) - f_\delta\|$ is decreasing, by Lemma 3, and $n_1 \le n_0$, one gets

$$\|F(V_{n_0+1}) - f_\delta\| \le \|F(V_{n_1+1}) - f_\delta\| \le C\delta. \tag{120}$$

From (42), (119), (120), the relation (104), and the definition $C_1 = 2C - 1$ (see (100)), one concludes that

$$\begin{aligned} \|F(u_{n_0+1}) - f_\delta\| \le& \frac{M_1 a_{n_0+1}^2}{\lambda} + C\delta \\ \le& \frac{M_1 \delta (C-1)}{\lambda \|y\|} + C\delta \le (2C-1)\delta = C_1 \delta. \end{aligned} \tag{121}$$

Thus, if

$$\|F(u_0) - f_\delta\| > C_1 \delta^\zeta, \quad 0 < \zeta \le 1,$$

then one concludes from (121) *that there exists* n_δ, $0 < n_\delta \le n_0 + 1$, *such that*

$$\|F(u_{n_\delta}) - f_\delta\| \le C_1 \delta^\zeta < \|F(u_n) - f_\delta\|, \quad 0 \le n < n_\delta, \tag{122}$$

for any given $\zeta \in (0, 1]$, *and any fixed* $C_1 > 1$.

Let us prove (97).

If $n > 0$ is fixed, then $u_{\delta,n}$ is a continuous function of f_δ. Denote

$$\tilde{u} := \tilde{u}_N = \lim_{\delta \to 0} u_{\delta, n_{m_j}}, \tag{123}$$

where

$$\lim_{j \to \infty} n_{m_j} = N.$$

From (123) and the continuity of F, one obtains:

$$\|F(\tilde{u}) - f_\delta\| = \lim_{j \to \infty} \|F(u_{n_{m_j}}) - f_\delta\| \le \lim_{\delta \to 0} C_1 \delta^\zeta = 0.$$

Thus, $\tilde{u}$ is a solution to the equation $F(u) = f$, and (97) is proved.

Let us prove (99) *assuming that* (98) *holds.*

From (96) and (119) with $n = n_\delta - 1$, and from (122), one gets

$$C_1 \delta^\zeta \le M_1 \frac{a_{n_\delta - 1}^2}{\lambda} + a_{n_\delta - 1} \|V_{n_\delta - 1}\| \le M_1 \frac{a_{n_\delta - 1}^2}{\lambda} + \|y\| a_{n_\delta - 1} + \delta.$$

If $\delta > 0$ is sufficiently small, then the above equation implies

$$\tilde{C}\delta^\zeta \le a_{n_\delta - 1}\Big(\frac{M_1 a_0}{\lambda} + \|y\|\Big), \quad \tilde{C} > 0,$$

where $\tilde{C} < C_1$ is a constant, and the inequality $a_{n_\delta - 1}^2 \le a_{n_\delta - 1} a_0$ was used. Therefore, by (40),

$$\lim_{\delta \to 0} \frac{\delta}{2a_{n_\delta}} \le \lim_{\delta \to 0} \frac{\delta}{a_{n_\delta - 1}} \le \lim_{\delta \to 0} \frac{\delta^{1-\zeta}}{\tilde{C}} \Big(\frac{M_1 a_0}{\lambda} + \|y\|\Big) = 0, \quad 0 < \zeta < 1. \tag{124}$$

In particular, for $\delta = \delta_m$, one gets

$$\lim_{\delta_m \to 0} \frac{\delta_m}{a_{n_m}} = 0. \tag{125}$$

From the triangle inequality and inequalities (15) and (115) one obtains

$$\begin{aligned} \|u_{n_m} - y\| &\le \|u_{n_m} - V_{n_m}\| + \|V_n - V_{n_m,0}\| + \|V_{n_m,0} - y\| \\ &\le \frac{a_{n_m}^2}{\lambda} + \frac{\delta_m}{a_{n_m}} + \|V_{n_m,0} - y\|. \end{aligned} \tag{126}$$

From (98), (125), inequality (126) and Lemma 1, one obtains (99). Theorem 19 is proved. $\Box$

4 Numerical experiments

Let us do a numerical experiment solving nonlinear equation (1) with

$$F(u) := B(u) + \frac{u^3}{6} := \int_0^1 e^{-|x-y|} u(y)dy + \frac{u^3}{6}, \quad f(x) := \frac{13}{6} - e^{-x} - \frac{e^x}{e}. \tag{127}$$

Such equation is a model nonlinear equation in Wiener-type filtering theory, see [18].

One can check that $u(x) \equiv 1$ solves the equation $F(u) = f$. The operator B is compact in $H = L^2[0,1]$. The operator $u \longmapsto u^3$ is defined on a dense subset D of of $L^2[0,1]$, for example, on $D := C[0,1]$. If $u, v \in D$, then

$$\langle u^3 - v^3, u - v \rangle = \int_0^1 (u^3 - v^3)(u - v)dx \ge 0.$$

Moreover,

$$e^{-|x|} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{e^{i\lambda x}}{1 + \lambda^2} d\lambda.$$

Therefore, $\langle B(u-v), u-v\rangle \geq 0$, so

$$\langle F(u-v), u-v\rangle \geq 0, \qquad \forall u, v \in D.$$

Note that D does not contain subsets, open in $H = L^2[0,1]$, i.e., it does not contain interior points of H. This is a reflection of the fact that the operator $G(u) = \frac{u^3}{6}$ is unbounded on any open subset of H. For example, in any ball $\|u\| \leq C$, $C = const > 0$, where $\|u\| := \|u\|_{L^2[0,1]}$, there is an element u such that $\|u^3\| = \infty$. As such an element one can take, for example, $u(x) = c_1 x^{-b}$, $\frac{1}{3} < b < \frac{1}{2}$. here $c_1 > 0$ is a constant chosen so that $\|u\| \leq C$. The operator $u \longmapsto F(u) = G(u) + B(u)$ is maximal monotone on $D_F := \{u : u \in H,\ F(u) \in H\}$ (see [1, p.102]), so that equation (8) is uniquely solvable for any $f_\delta \in H$.

The Fréchet derivative of F is:

$$F'(u)h = \frac{u^2 h}{2} + \int_0^1 e^{-|x-y|} h(y)dy. \tag{128}$$

If $u(x)$ vanishes on a set of positive Lebesgue's measure, then $F'(u)$ is obviously not boundedly invertible. If $u \in C[0,1]$ vanishes even at one point x_0, then $F'(u)$ is not boundedly invertible in H.

Let us use the iterative process (93):

$$\begin{aligned} u_{n+1} &= u_n - \alpha_n \big(F'(u_n)^* + a_n I\big)(F(u_n) + a_n u_n - f_\delta), \\ u_0 &= 0. \end{aligned} \tag{129}$$

We stop iterations at $n := n_\delta$ such that the following inequality holds

$$\|F(u_{n_\delta}) - f_\delta\| < C\delta^\zeta, \quad \|F(u_n) - f_\delta\| \geq C\delta^\zeta, \quad n < n_\delta, \quad C > 1, \quad \zeta \in (0,1). \tag{130}$$

Integrals of the form $\int_0^1 e^{-|x-y|} h(y)dy$ in (127) and (128) are computed by using the trapezoidal rule. The noisy function used in the test is

$$f_\delta(x) = f(x) + \kappa f_{noise}(x), \quad \kappa > 0.$$

The noise level δ and the relative noise level are determined by

$$\delta = \kappa \|f_{noise}\|, \quad \delta_{rel} := \frac{\delta}{\|f\|}.$$

In the test, κ is computed in such a way that the relative noise level δ_{rel} equals to some desired value, i.e.,

$$\kappa = \frac{\delta}{\|f_{noise}\|} = \frac{\delta_{rel}\|f\|}{\|f_{noise}\|}.$$

We have used the relative noise level as an input parameter in the test.

The version of DSM, developed in this paper and denoted by DSMG, is compared with the version of DSM in [3], denoted by DSMN. Indeed, the DSMN is the following iterative scheme

$$u_{n+1} = u_n - A_n^{-1}\big(F'(u_n) + a_n u_n - f_\delta\big), \quad u_0 = u_0, \qquad n \geq 0, \tag{131}$$

where $a_n = \frac{a_0}{1+n}$. This iterative scheme is used with a stopping time n_δ defined by (96). The existence of this stopping time and the convergence of the method is proved in [3].

As we have proved, the DSMG converges when $a_n = \frac{a_0}{(1+n)^b}$, $b \in (0, \frac{1}{4}]$, and a_0 is sufficiently large. However, in practice, if we choose a_0 too large then the method will use too many iterations before reaching the stopping time n_δ in (130). This means that the computation time is large. Since

$$\|F(V_\delta) - f_\delta\| = a(t)\|V_\delta\|,$$

and $\|V_\delta(t_\delta) - u_\delta(t_\delta)\| = O(a(t_\delta))$, we have

$$C\delta^\zeta = \|F(u_\delta(t_\delta)) - f_\delta\| \sim a(t_\delta).$$

Thus, we choose

$$a_0 = C_0\delta^\zeta, \qquad C_0 > 0.$$

The parameter a_0 used in the DSMN is also chosen by this formula.

In all figures, the x-axis represents the variable x. In all figures, by *DSMG* we denote the numerical solutions obtained by the DSMG, by *DSMN* we denote solutions by the DSMN and by *exact* we denote the exact solution.

In experiments, we found that the DSMG works well with $a_0 = C_0\delta^\zeta$, $C_0 \in [0.2, 1]$. Indeed, in the test the DSMG is implemented with $a_n := C_0\frac{\delta^{0.99}}{(n+1)^{0.25}}$, $C_0 = 0.5$ while the DSMN is implemented with $a_n := C_0\frac{\delta^{0.99}}{(n+1)}$, $C_0 = 1$. For $C_0 > 1$ the convergence rate of DSMG is much slower while the DSMN still works well if $C_0 \in [1, 4]$.

Figure 1 plots the solutions using relative noise levels $\delta = 0.01$ and $\delta = 0.001$. The exact solution used in these experiments is $u = 1$. In the test the DSMG is implemented with $\alpha_n = 1$, $C = 1.01$, $\zeta = 0.99$ and $\alpha_n = 1$, $\forall n \geq 0$. The number of iterations of the DSMG for $\delta = 0.01$ and $\delta = 0.001$ were 49 and 50 while the number of iteration for the DSMN are 9 and 9, respectively. The number of node points used in computing integrals in (127) and (128) was $N = 100$. The noise function f_{noise} in this experiment is a vector with random entries normally distributed of mean 0 and variant 1. Figure 1 shows that the solutions by the DSMN and DSMG are nearly the same in this figure.

Figure 2 presents the numerical results when $N = 100$ with $\delta = 0.01$ $u(x) = \sin(2\pi x)$, $x \in [0, 1]$ (left) and with $\delta = 0.01$, $u(x) = \sin(\pi x)$, $x \in [0, 1]$ (right). In these cases, the DSMN took 11 and 7 iterations to give the numerical solutions while the DSMG took 512 and 94 iterations for $u(x) = \sin(2\pi x)$ and $u(x) = \sin(\pi x)$, respectively. Figure 2 show that the numerical results of the DSMG are better than those of the DSMN.

Numerical experiments agree with the theory that the convergence rate of the DSMG is slower than that of the DSMN. It is because the rate of decaying of the sequence $\{\frac{1}{(1+n)^{\frac{1}{4}}}\}_{n=1}^\infty$ is much slower than that of the sequence $\{\frac{1}{1+n}\}_{n=1}^\infty$. However, if the cost for evaluating F and F' are not counted then the cost of computation at one iteration of the DSMG is of $O(N^2)$ while that of the DSMN in one iteration of the DSMN is of $O(N^3)$. Here N is the number of the nodal points. Thus, for large scale problems, the DSMG might be an alternative to the DSMN. Also, as it is showed in Figure 2, the DSMG might yield solutions with better accuracy.

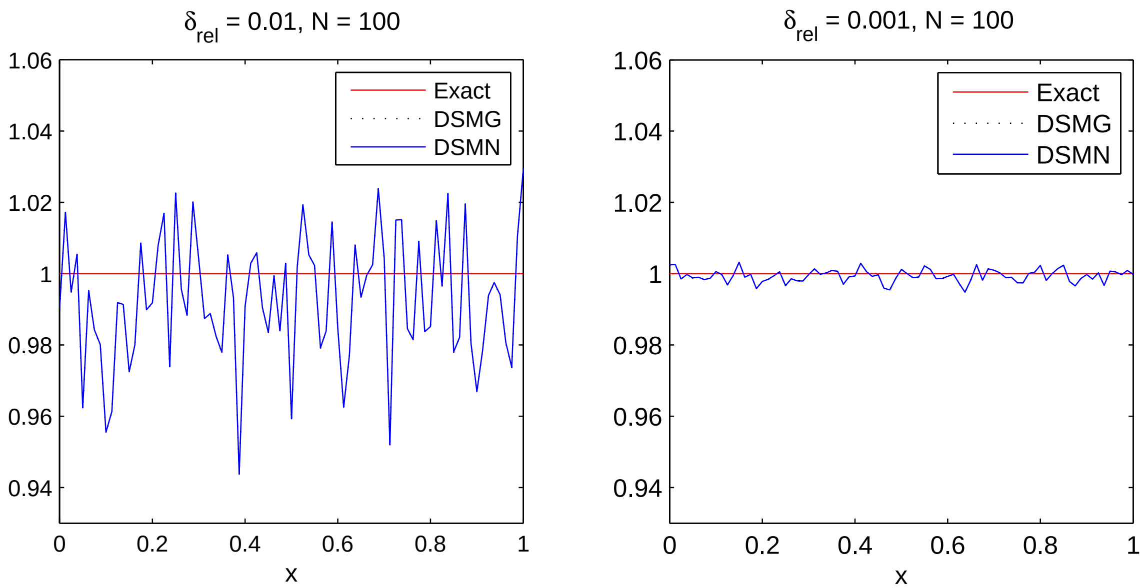


Figure 1: Plots of solutions obtained by the DSMN and DSMG when $N = 100$, $u = 1$, $x \in [0, 1]$, $\delta_{rel} = 0.01$ (left) and $N = 100$, $u = 1$, $x \in [0, 1]$, $\delta_{rel} = 0.001$ (right).

Experiments show that the DSMN still works with $a_n = \frac{a_0}{(1+n)^b}$ for $\frac{1}{4} \le b \le 1$. So in practice, one might use faster decaying sequence a_n to reduce the time of computation.

From the numerical results we conclude that the proposed stopping rule yields good results in this problem.

References

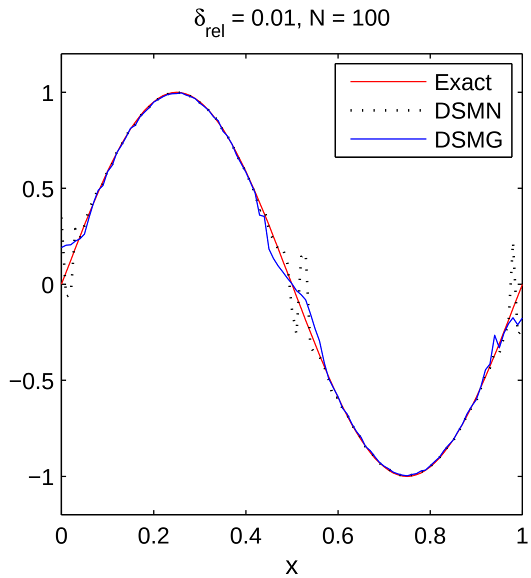
[1] K. Deimling, Nonlinear functional analysis, Springer Verlag, Berlin, 1985.

[2] N. S. Hoang and A. G. Ramm, Solving ill-conditioned linear algebraic systems by the dynamical systems method, Inverse Probl. in Sci. and Engineering, 16, N5, (2008), 617-630.

[3] N. S. Hoang and A. G. Ramm, An iterative scheme for solving nonlinear equations with monotone operators. (submitted)

[4] V. Ivanov, V. Tanana and V. Vasin, Theory of ill-posed problems, VSP, Utrecht, 2002.

[5] J. L. Lions, Quelques methodes de resolution des problemes aux limites non lineaires, Dunod, Gauthier-Villars, Paris, 1969.



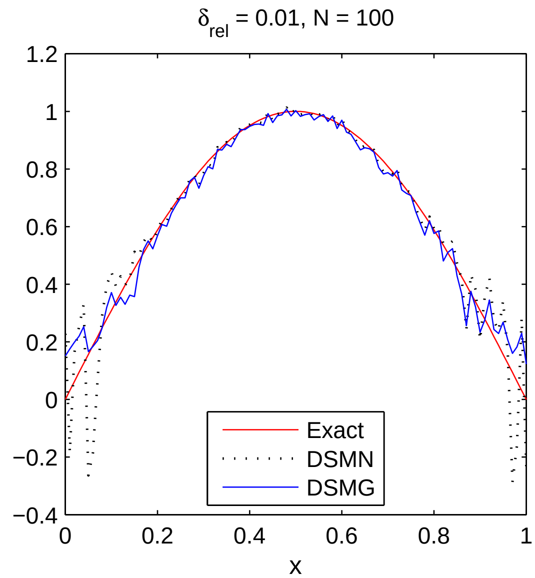


Figure 2: Plots of solutions obtained by the DSMN and DSMG when $N = 100$, $u(x) = \sin(2\pi x)$, $x \in [0, 1]$, $\delta_{rel} = 0.01$ (left) and $N = 100$, $u(x) = \sin(\pi x)$, $x \in [0, 1]$, $\delta_{rel} = 0.01$ (right).

[6] V. A. Morozov, Methods of solving incorrectly posed problems, Springer Verlag, New York, 1984.

[7] D. Pascali and S. Sburlan, Nonlinear Mappings of Monotone Type, Noordhoff, Leyden, 1978.

[8] A. G. Ramm, Theory and applications of some new classes of integral equations, Springer-Verlag, New York, 1980.

[9] A. G. Ramm, Stationary regimes in passive nonlinear networks, in the book "Nonlinear Electromagnetics", Ed. P.Uslenghi, Acad. Press, New York, 1980, pp. 263-302.

[10] A. G. Ramm, Dynamical systems method for solving operator equations, Elsevier, Amsterdam, 2007.

[11] A. G. Ramm, Global convergence for ill-posed equations with monotone operators: the dynamical systems method, J. Phys A, 36, (2003), L249-L254.

[12] A. G. Ramm, Dynamical systems method for solving nonlinear operator equations, International Jour. of Applied Math. Sci., 1, N1, (2004), 97-110.

[13] A. G. Ramm, Dynamical systems method for solving operator equations, Communic. in Nonlinear Sci. and Numer. Simulation, 9, N2, (2004), 383-402.

[14] A. G. Ramm, DSM for ill-posed equations with monotone operators, Comm. in Nonlinear Sci. and Numer. Simulation, 10, N8, (2005),935-940.

[15] A. G. Ramm, Discrepancy principle for the dynamical systems method, Communic. in Nonlinear Sci. and Numer. Simulation, 10, N1, (2005), 95-101

[16] A. G. Ramm, Dynamical systems method (DSM) and nonlinear problems, in the book: Spectral Theory and Nonlinear Analysis, World Scientific Publishers, Singapore, 2005, 201-228. (ed J. Lopez-Gomez).

[17] A. G. Ramm, Dynamical systems method (DSM) for unbounded operators, Proc. Amer. Math. Soc., 134, N4, (2006), 1059-1063.

[18] A. G. Ramm, Random fields estimation, World Sci. Publishers, Singapore, 2005.

[19] A. G. Ramm, Iterative solution of linear equations with unbounded operators, J. Math. Anal. Appl., 1338-1346.

[20] A. G. Ramm, On unbounded operators and applications, Appl.Math. Lett., 21, (2008), 377-382.

[21] I.V. Skrypnik, Methods for Analysis of Nonlinear Elliptic Boundary Value Problems, American Mathematical Society, Providence, RI, 1994.

[22] U. Tautenhahn, On the asymptotical regularization method for nonlinear ill-posed problems, Inverse Problems, 10, (1994), pp.1405-1418.

[23] U. Tautenhahn, On the method of Lavrentiev regularization for nonlinear ill-posed problems, Inverse Probl., 18, (2002), 191-207.

[24] M.M. Vainberg, Variational methods and method of monotone operators in the theory of nonlinear equations, Wiley, London, 1973.

[25] E. Zeidler, Nonlinear functional analysis, Springer, New York, 1985.